Findings from the DIL Interviews: Data Curation and Reuse

Skills in this competency may include:

- Recognizes that data may have value beyond the original purpose, to validate research, or for use by others.
- Distinguishes which elements of a data set are likely to have future value for self and for others.
- Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven e-research.
- Recognizes that data must be prepared for its eventual curation at its creation and throughout its lifecycle.
- Articulates the planning and activities needed to enable data curation, both generally and within his/her local practice.
- Understands how to cite data as well as how to make his/her data citable.

Average Ranking of Importance (5=essential): Faculty=4.25, Students = 4.06

Faculty responses:

Faculty view this area as important, but that the skills are lacking by both students and the researchers themselves. In fact, several commented that this idea of data reuse is just beginning to take hold. One faculty commented that the entire research lab needs a better understanding of who would benefit from data curation. Another felt that students generally won't have to concern themselves with these skills as it is up to the researcher.

Faculty also see this area as very important because data cannot simply be recreated over the course of extended experiments. Not all data is viable for curation however, as one faculty noted code that is not standardized is undesirable and doesn't promote future research. Even so, faculty commented that the academic culture may place less emphasis on complete functionality for the public, focusing more on the research itself.

Some faculty have deposited data in a repository, some have not. The data is posted online, deposited code into SourceForge and Google Code. However, getting the software in a format where it is ready to be shared with the rest of the world may be difficult.

Student responses:

Students recognize which specific data sets (raw vs. processed vs. published) will be most valuable to save, but may not immediately understand the potential value for reuse in the data they create. The practices and skills to preserve digital information is even less understood. For example, one student believed that individuals in the lab were taking the necessary steps to prepare the generated data for eventual reuse, but was unsure of "exactly what they're doing."